Importance Sampling and Randomized QMC

Art B. Owen & Sifan Liu
Stanford University

An integral

$$\int_{\mathbb{R}^d} \sum_{i=1}^{N} \frac{y_i x_{ij} z_j}{1+e^{y_i \boldsymbol{x}_i^\mathsf{T} \beta(\boldsymbol{z})}} \frac{e^{-\boldsymbol{z}^\mathsf{T}\boldsymbol{z}/2}}{(2\pi)^{d/2}}\,\mathrm{d}\boldsymbol{z} \qquad \beta_k = \mu_k + \sigma_k z_k$$

Given data

$N \qquad y_i \in \{-1,1\} \qquad \boldsymbol{x}_i \in \mathbb{R}^d \qquad \mu \in \mathbb{R}^d \qquad \sigma \in (0,\infty)^d$

Origin

Variational Bayes approximation to a posterior distribution in logistic regression

Gradient from the ***evidence lower bound*** (ELBO)

for $\beta \sim \mathcal{N}(\mu, \text{diag}(\sigma^2))$

From Sifan Liu & O (2021)

Thursday Zoom 5 20:00

Unbounded

Due to z_j

RQMC

Here it means scrambled digital nets

mostly Sobol' nets

scrambling could be

nested uniform O (1995), or

affine Matousek (1998)

Our goal

Better results for unbounded integrands

Ideally $O(n^{-3/2+\epsilon})$

Hard for large d

Even $O(n^{-1+\epsilon})$

Would be great.

NB: Unbounded integrands are not BVHK

Common rates

Akeson & Lehoczky (2000) Mortgage backed securities

Estimated rates in 0.5 to 1.0

L'Ecuyer (2009) Options

Variance reduction factors (VRF) increase with n

Empirically VRF$=o(n)$ so RMSE not as good as $O(n^{-1})$

This talk

We look into importance sampling (IS)

The obvious approach brings a curse of dimension

$O(n^{-3/2+\epsilon})$ would happen at infeasible n

We introduce a ***minimal*** importance sampler

We are still far from reliable $O(n^{-1+\epsilon})$

Goal continued

All Lipschitz functions of $\boldsymbol{z} \sim \mathcal{N}(0, I)$

Too hard. Curse of dimension Curbera (2000)

More reasonable

Polynomial or exponential growth

$$f(\boldsymbol{z}) = O(e^{\kappa\|\boldsymbol{z}\|}) \quad \text{or} \quad O(\|\boldsymbol{z}\|^k) \quad \text{or} \quad O(\|\boldsymbol{z}\|)$$

Plus adequate smoothness

Approximate ridge functions

Our long term goal

$$f(\boldsymbol{z}) = g(\Theta^\mathsf{T}\boldsymbol{z}) + \varepsilon(\boldsymbol{z})$$

$\Theta \in \mathbb{R}^{d\times r}$ $\qquad r \ll d$ $\qquad g(\cdot)$ Lipschitz

Gaussian notation

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\Bigl(-\frac{z^2}{2}\Bigr)$$

$$\Phi(z) = \int_{-\infty}^{z} \varphi(x)\,\mathrm{d}x$$

$$x \sim \mathbf{U}(0,1) \implies z = \Phi^{-1}(x) \sim \mathcal{N}(0,1)$$

Toy functions

$$f(\boldsymbol{z}) = z_1$$

Linear in one input

$\theta^{\mathsf{T}}\boldsymbol{z}$ for known θ

Anything that fails for this one cannot serve

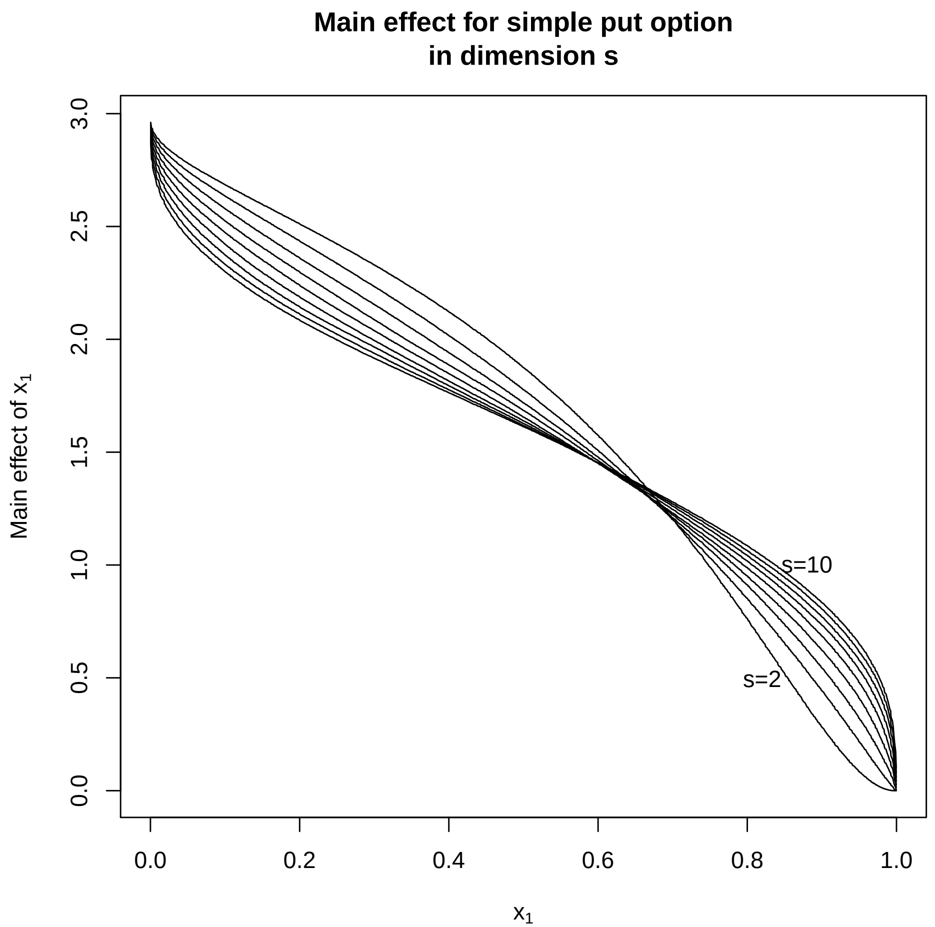
Simple put option

$$f_{\mathrm{put}}(\boldsymbol{z}) = \Bigl(K - e^{\sum_{j=1}^{d} z_j}\Bigr)_+$$

Bounded but has similar challenge as call option

Put main effects

$$\mu + f_{\{j\}}(z_j) = \int_{\mathbb{R}^{d-1}} f_{\rm put}(\boldsymbol{z}) \prod_{k \neq j} \varphi(z_k)\, \mathrm{d}z_k \qquad \text{versus } x_j = \Phi(z_j)$$



Scrambled net properties

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \hat\mu = \hat\mu_n = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)$$

If	Then	
$f \in L^1$	$\mathbb{E}(\hat\mu) = \mu$	
$f \in L^{1+\delta}$	$\Pr\bigl(\lim_{n\to\infty} \hat\mu_n = \mu\bigr) = 1$	O & Rudolf (2020)
$f \in L^2$	$\mathrm{RMSE}(\hat\mu) = o(n^{-1/2})$	
$f \in L^2$	$\mathrm{RMSE}(\hat\mu) \leqslant \Gamma^{1/2}\sigma n^{-1/2}$	some $\Gamma < \infty$
$f \in \mathrm{BVHK}$	$\mathrm{RMSE}(\hat\mu) = O(n^{-1+\epsilon})$	
f "smooth"	$\mathrm{RMSE}(\hat\mu) = O(n^{-3/2+\epsilon})$	

Plain MC has $\mathrm{RMSE}(\hat\mu) = \sigma n^{-1/2} \quad \sigma^2 = \mathrm{Var}(f(\boldsymbol{x}))$.

Pan & O (2021) $\Gamma \leqslant 2^{t+d-1}$ for Sobol'

Smooth

Suffices to have

$$\partial^u f = \prod_{j \in u} \frac{\partial}{\partial x_j} f(\boldsymbol{x}) \quad \text{continuous on } [0,1]^d \qquad \text{O (2008)}$$

for all $u \subseteq 1{:}d \equiv \{1, 2, \dots, d\}$

Sharpest conditions

Yue & Mao (1999)

Generalized Lipschitz condition

Cannot hold for unbounded f

Importance sampling

$$\mu = \mathbb{E}_p(f(\boldsymbol{z})) = \int f(\boldsymbol{z})p(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

Ordinary IS

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(\boldsymbol{z}_i)p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} \qquad \boldsymbol{z}_i \sim q$$

Self-normalized IS

Works with unnormalized p, q

If time permits, I'll say why it should make little difference

Some literature

Chelson (1976)

First use

Dick & Aistleitner (2014)

repaired Koksma-Hlawka

Hörmann & Leydold (2005,2007)

Use heavy tailed q, find empirical improvements

Chopin & Ridgeway (2017)

RQMC+IS more effective than MCMC on some Bayesian inference problems

Zhang, Wang & He (2021)

QMC+IS for finance; optimal drift IS & Laplace IS

For reducing effective dimension

Glasserman, Heidelberger & Shahabuddin (1999)

stratified sampling in optimal direction

Transformations

$$\boldsymbol{x} \sim \mathbf{U}(0,1)^d \qquad \tau(\boldsymbol{x}) \sim p \equiv \mathcal{N}(0,I) \qquad \tau(\boldsymbol{x}) = \Phi^{-1}(\boldsymbol{x})$$

Sample $\boldsymbol{z} = \phi(\tau(\boldsymbol{x}))$ to get $\boldsymbol{z} \sim q$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(\boldsymbol{z}_i)p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}_i)$$

$$\begin{aligned} g(\boldsymbol{x}) &= \frac{fp}{q}\circ\phi\circ\tau(\boldsymbol{x}) \\ &= (f\times w)\circ\phi\circ\tau(\boldsymbol{x}) \end{aligned}$$

$$w(\cdot) = \frac{p(\cdot)}{q(\cdot)}$$

Partial derivatives

For ***componentwise*** ϕ and τ and $fp/q = fw$

$$\partial^u((f\times w)\circ\phi\circ\tau)(\boldsymbol{x}) = \sum_{v\subseteq u}\partial^v(f\circ\phi\circ\tau(\boldsymbol{x}))\times\partial^{u-v}(w\circ\phi\circ\tau(\boldsymbol{x}))$$

$$\partial^u(f\circ\phi\circ\tau)(\boldsymbol{x}) = (\partial^v f)\bigl(\phi\circ\tau(\boldsymbol{x})\bigr)\prod_{j\in u}\phi'(\tau'(x_j))$$

These come from simplified Faa di Bruno theorems

Special case

$$p = \mathcal{N}(0, I) \qquad q = \mathcal{N}(0, \lambda^2 I)$$

$$\boldsymbol{z} = \phi(\tau(\boldsymbol{x})) = \lambda \times \tau(\boldsymbol{x}) \qquad \lambda > 1$$

$$\begin{aligned}\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{z}_i)\frac{p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} \\ &= \cdots \\ &= \frac{\lambda^d}{n}\sum_{i=1}^{n} f(\boldsymbol{z}_i)\exp\Bigl(\Bigl(\frac{1}{2\lambda^2}-\frac{1}{2}\Bigr)\boldsymbol{z}_i^{\mathsf{T}}\boldsymbol{z}_i\Bigr)\end{aligned}$$

Taking $\lambda > 1$ will make a bounded integrand if $|f(\boldsymbol{z})| \leqslant \exp(\kappa\|\boldsymbol{z}\|)$

Derivative for $d = 1$

$$p(z) = \varphi(z) \qquad q(z) = \frac{1}{\lambda}\varphi\Bigl(\frac{z}{\lambda}\Bigr)$$

$$\frac{\partial}{\partial x}\frac{f(z)p(z)}{q(z)} = \sqrt{2\pi}\exp\Bigl(\Bigl(\frac{1}{\lambda^2}-\frac{1}{2}\Bigr)z^2\Bigr) \times \Bigl(f'(z) + zf(z)(\lambda^{-3}-1)\Bigr)$$

where $z = z(x) = \lambda \times \Phi^{-1}(x)$

Upshot

We need $\lambda > \sqrt{2}$ for smoothness

Still need that for $d > 1$

More generally

$$-\frac{1}{2}z^2 + \phi^{-1}(z)^2 \leqslant B < \infty$$

$$-\frac{1}{2}\phi(z)^2 + z^2 \leqslant B < \infty$$

Product

$$w(\boldsymbol{z}) = \prod_{j=1}^{d} \frac{p_j(z_j)}{q_j(z_j)}$$

$$\begin{aligned}
\mathrm{Var}_q(w(\boldsymbol{z})) &= \mathbb{E}_{q_1}\Bigl(\Bigl(\frac{p_1(z_1)}{q_1(z_1)}\Bigr)^2\Bigr)^d - \mathbb{E}_{q_1}\Bigl(\frac{p_1(z_1)}{q_1(z_1)}\Bigr)^d \\
&= \mathbb{E}_{q_1}\Bigl(\Bigl(\frac{p_1(z_1)}{q_1(z_1)}\Bigr)^2\Bigr)^d - 1 \\
&> 1.63^d - 1
\end{aligned}$$

for $\lambda > \sqrt{2}$

Large norm and nowhere near low effective dimension

Mean dimension $\approx 0.39d$

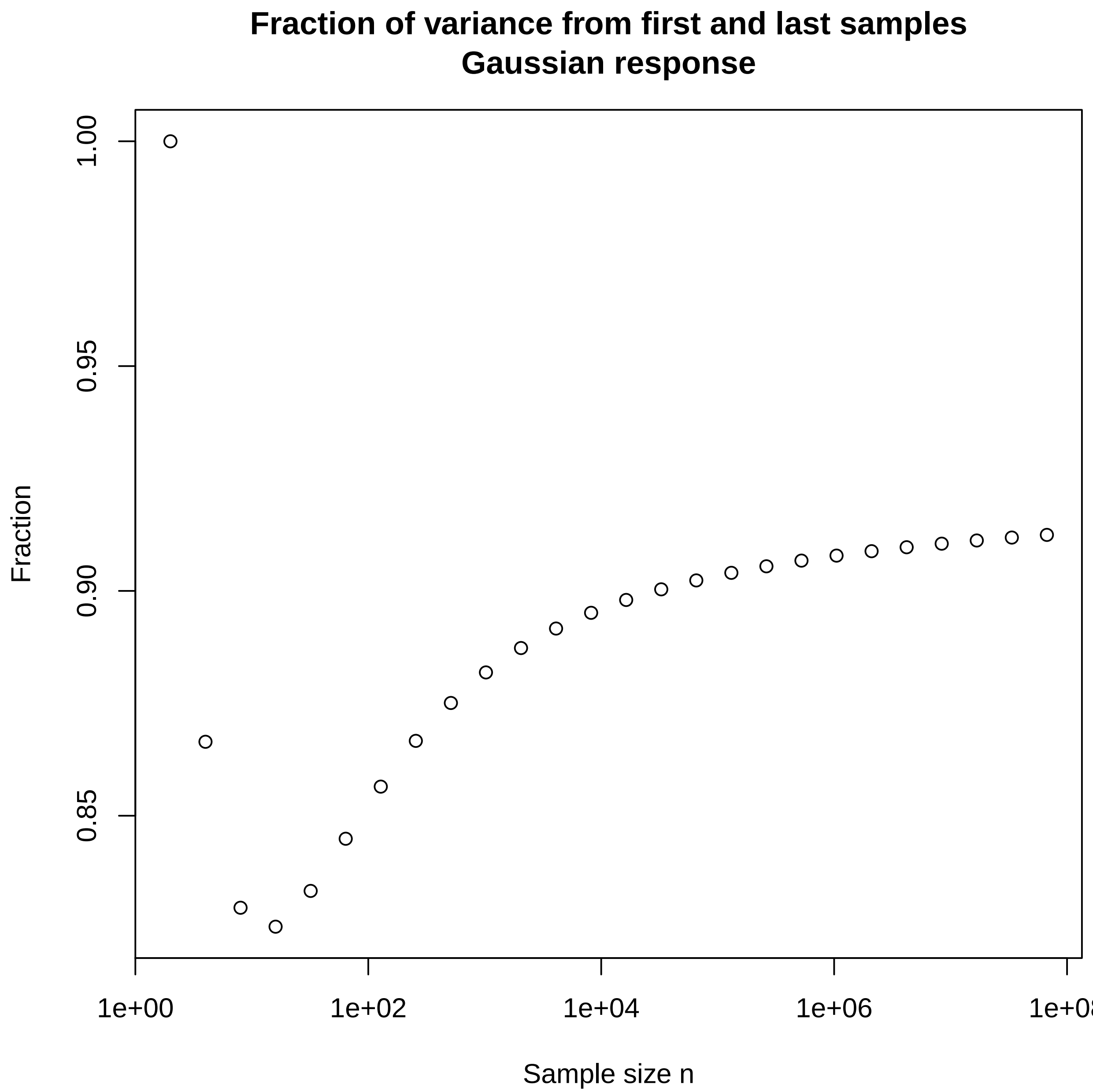
Similar thing happens for periodization transformations

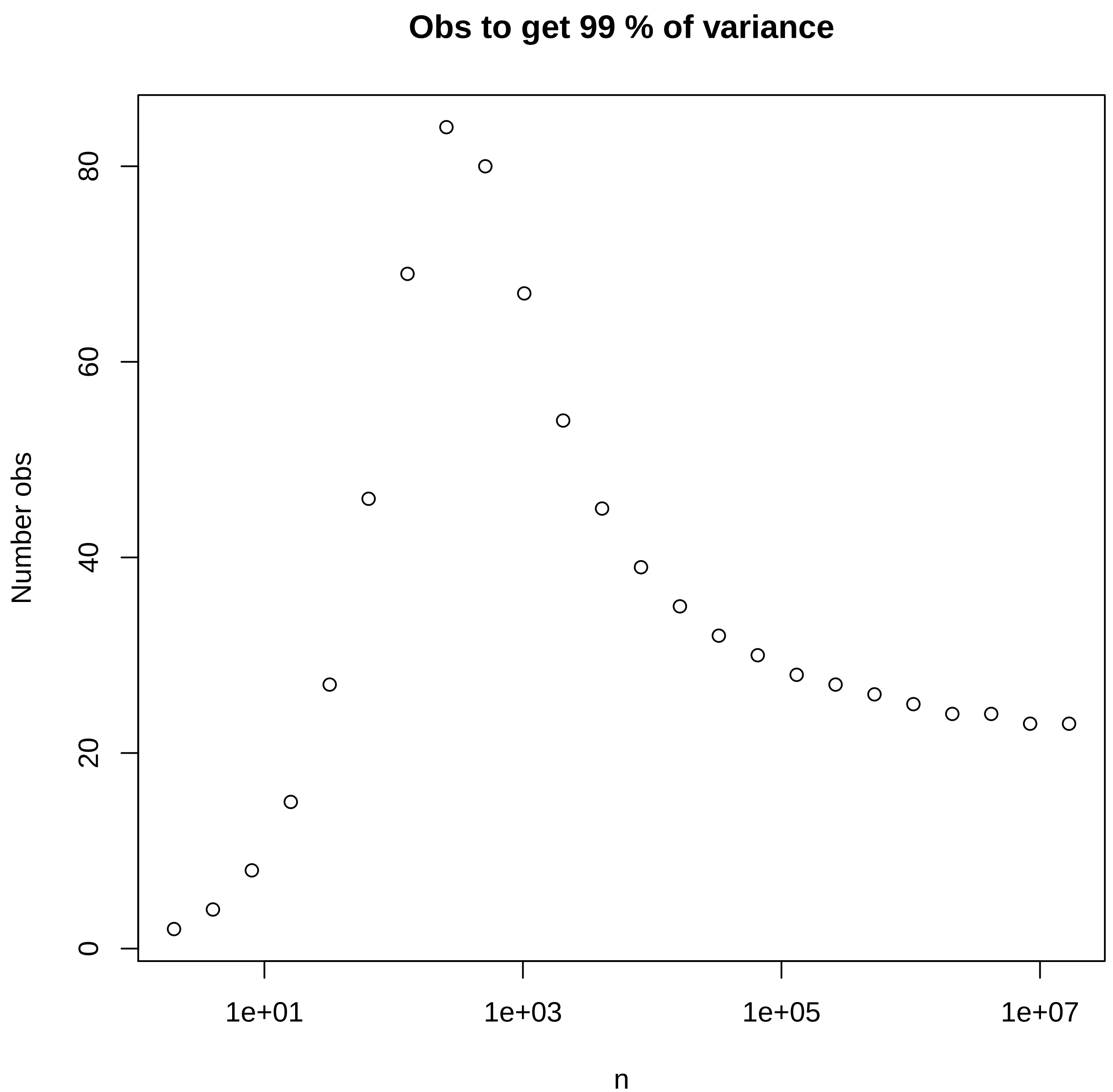
Toy $f(\boldsymbol{z}) = z_1$

For nested uniform scrambling $x_i \overset{\text{ind}}{\sim} \mathbf{U}\bigl(\frac{i-1}{n}, \frac{i}{n}\bigr)$

$$\underbrace{\Phi^{-1}\Bigl(\frac{i-1}{n}\Bigr)}_{\alpha_i} \leqslant z_{i1} \leqslant \underbrace{\Phi^{-1}\Bigl(\frac{i}{n}\Bigr)}_{\beta_i}$$

$$\begin{aligned}
\mathrm{Var}\Bigl(\frac{1}{n}\sum_{i=1}^n z_{i1}\Bigr) &= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\bigl(\Phi^{-1}(x_{i1})\bigr) \\
&= \frac{1}{n^2}\sum_{i=1}^n \Biggl[1 + \frac{\alpha_i\varphi(\alpha_i) - \beta_i\varphi(\beta_i)}{\Phi(\beta_i) - \Phi(\alpha_i)} - \Bigl(\frac{\varphi(\alpha_i) - \varphi(\beta_i)}{\Phi(\beta_i) - \Phi(\alpha_i)}\Bigr)^2\Biggr]
\end{aligned}$$


Fraction of variance from first and last samples
Gaussian response
Fraction
1.00
0.95
0.90
0.85
1e+00
1e+02
1e+04
1e+06
1e+08
Sample size n


Obs to get 99 % of variance
Number obs
0
20
40
60
80
1e+01
1e+03
1e+05
1e+07
n

Put option functions

Main effects also dominated by first and last points

Importance sample 2 of n points

$$\underbrace{\Phi^{-1}\Bigl(\frac{i-1}{n}\Bigr)}_{\alpha_i} \leqslant z_i \leqslant \underbrace{\Phi^{-1}\Bigl(\frac{i}{n}\Bigr)}_{\beta_i}$$

Nominal

$$p_{i,n}(z) = \varphi(z)1_{\alpha_i \leqslant z \leqslant \beta_i}$$

Sampling

$q_n(z_n) \propto z\varphi(z)$ on $z > \Phi^{-1}(1-1/n)$

exact for $f(z) = z$

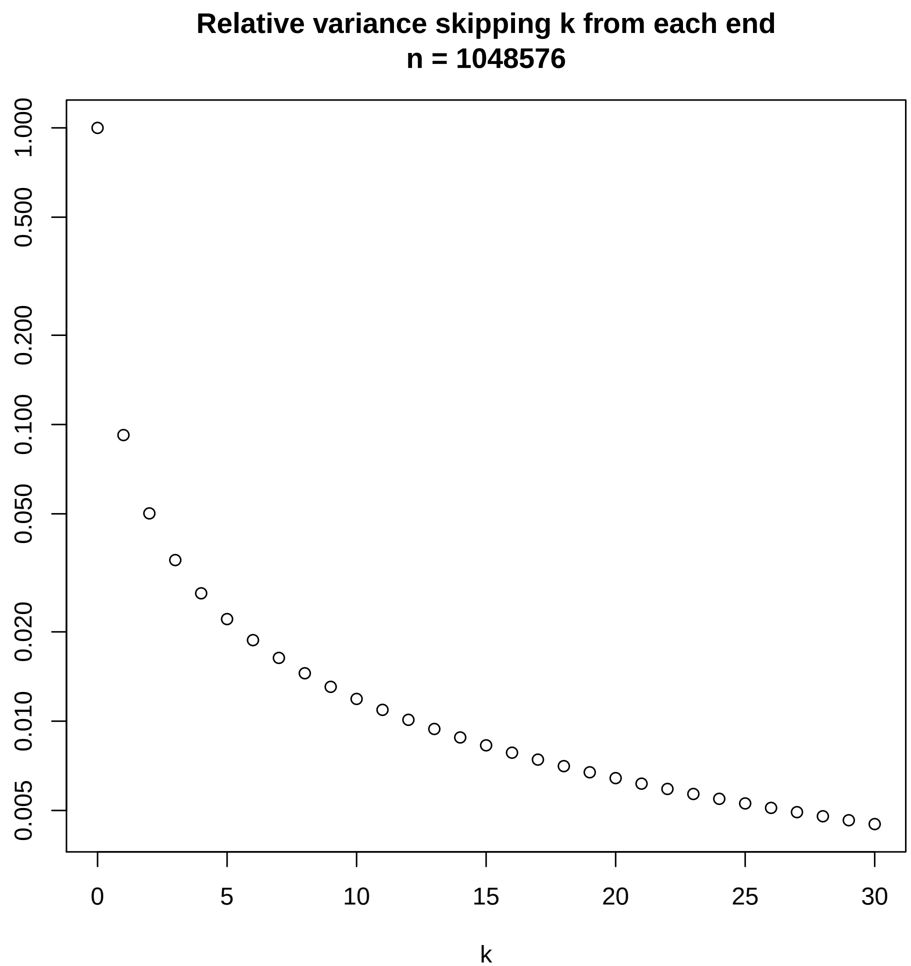
similar for $q_1(z_1)$

but not $\sqrt{2}$ times bigger

Rest are nominal

$$f(x) = x$$

Importance sample $2k$ pts exactly



z_1 here is $\theta^{\mathsf{T}}\boldsymbol{z}$ in a ridge

Not easy to extend to other f

Variance of p/q

$$A_n \equiv \Phi^{-1}(1-1/n) \approx \sqrt{2\log(n)}$$

Asymptotic trouble

$$\begin{aligned}\mathbb{E}_q\Bigl(\Bigl(\frac{p(z)}{q(z)}\Bigr)^2\Bigr) &= \int_{-\infty}^{\infty}\frac{p(z)^2}{q(z)}\,\mathrm{d}z\\ &\leqslant 1-2\Phi(-A_n)+\frac{\varphi(A_n)^2}{\Phi(-A_n)}\log\Bigl(1+\frac{2}{A_n^2}\Bigr)\\ &\lesssim 1+\frac{3}{n\log(n)}\end{aligned}$$

Empirically

$$\mathbb{E}_q\Bigl(\Bigl(\frac{p(z)}{q(z)}\Bigr)^2\Bigr) \leqslant \frac{1.08}{n\log(n)} \quad n=2^m \quad 2\leqslant m\leqslant 40$$

d dim variance

$$\Bigl(1+\frac{c}{n\log(n)}\Bigr)^d - 1$$

Thanks

1) Sifan Liu, co-author

2) NSF IIS-1837931

3) Andreas Neuenkirch and the Local Organizers

Importance sampling

$$\mu = \mathbb{E}_p(f(\boldsymbol{z})) = \int f(\boldsymbol{z})p(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

Ordinary IS

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(\boldsymbol{z}_i)p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} \qquad \boldsymbol{z}_i \sim q$$

Self-normalized IS

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(\boldsymbol{z}_i)p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} \Big/ \frac{1}{n}\sum_{i=1}^{n}\frac{p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)} \qquad \boldsymbol{z}_i \sim q$$

$$\doteq \mu + \frac{1}{n}\sum_{i=1}^{n}\frac{(f(\boldsymbol{z}_i)-\mu)p(\boldsymbol{z}_i)}{q(\boldsymbol{z}_i)}$$

After Taylor expansion (delta method)

Like using $\mathbb{E}_q(p/q) = 1$ as a control variate